Conjoint Routing and Resource Allocation in OFDMA-based D2D Wireless Networks

Rozita Rashtchi, Ramy H. Gohary, and Halim Yanikomeroglu

***Abstract*—In this paper, we develop a highly efficient two-tier technique for jointly optimizing the routes, the subcarrier schedules, the time-shares and the power allocations in device-to-device communication networks with thousands of randomly dropped wireless nodes. The network is first divided into a set of non-overlapping sub-networks, each with its own regional controller. The role of such a controller is to optimize the sub-network within its region and to act as an interface between nodes communicating across regions. The first tier of the proposed technique uses a novel approach for splitting a set of highly non-convex constraints into effectively two sets of convex ones and optimization proceeds by using two loops: an outer loop for iterating between the power allocations and the subcarrier schedules, and an inner loop for iterating between the two sides of the split constraints. In the second tier, a technique analogous to the one used in the first tier is applied to the network composed of the regional controllers. Optimization in this tier is performed by a global controller. The proposed technique is capable of efficiently optimizing networks with tens of thousands of nodes and with significantly better performance than existing joint design techniques, which can only optimize networks with a few tens of nodes.**

I. Introduction

The soon-to-be-standardized fifth-generation (5G) wireless networks will support device-to-device (D2D) communications in order to provide ubiquitous and reliable high-rate connectivity between a massive number of wireless communication devices [1], [2]. A key ingredient that will enable D2D communication systems to make better use of the available spectral resources, to increase system capacity, and to expand coverage is to use either fixed or device relaying techniques. Fixed relaying, which involves the deployment of low-power base stations (BSs) to assist cellular communications, has been extensively studied in the literature, e.g., [3], [4] and it has already been included in the fourth-generation (4G) Long Term Evolution (LTE)-Advanced standard. As the number of devices with higher demands increases in cellular networks, more relays must be deployed. This makes the network denser and hence increases the negative effects of interference. Several techniques such as inter-cell interference coordination [5] and coordinated beamforming [6] have been proposed to mitigate interference in 4G networks. In contrast, in future 5G networks it is desirable to exploit the network density to route data through a massive mesh network [7]. Such a potential is offered by D2D communications, wherein two devices are allowed to communicate in the licensed cellular bandwidth possibly without the involvement of the BS [8], see Figure 1. This is in contrast with conventional cellular architecture in which nodes communicate only with their BSs.

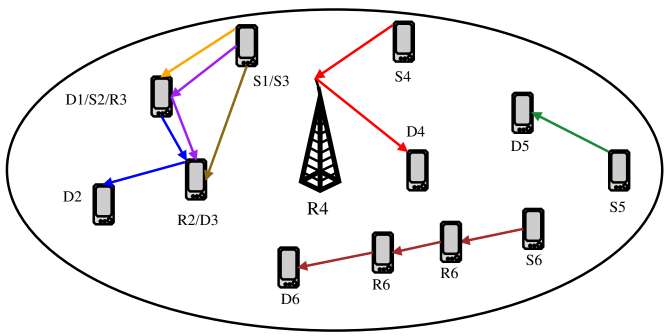


Fig. 1: A D2D communication scenario. Sources, destinations and relays are identified by S, D and R, respectively.

Emerging D2D networks are envisioned to use orthogonal frequency-division multiple access (OFDMA) as their air interface. This is mainly due to its simplicity and immunity to inter-symbol interference [9], in addition to its design flexibility and ability to achieve high spectral efficiency by scheduling the subcarriers based on the channel conditions of the users [10].

A main feature of D2D communications is the large number of inexpensive low-power devices competing for a scarce pool of radio resources. The number and versatility of services offered by these devices renders efficient utilization of the radio resources rather imperative, leaving little room for wasteful designs that do not benefit from the topological and propagation conditions of the network. In particular, efficient utilization of resources must take into consideration the network conditions when making decisions pertaining to routing, scheduling and power allocations. Although optimizing these aspects in isolation simplifies the design, it may result in wasting valuable resources that could otherwise be used to increase the network utility [11].

Joint optimization of routing, scheduling and power allocations in networks with a large number of D2D devices invokes several difficulties. For instance, the basic problem of optimizing power allocations in OFDMA networks is NP-hard [12] and even with fixed power allocations, the inherent combinatorial nature of the scheduling problem often renders the problem intractable. As such, joint optimization of scheduling and power allocation along with routing is usually computationally prohibitive. Several attempts for performing

R. Rashtchi, R.H. Gohary and H. Yanikomeroglu are with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada, {rozita, gohary, halim}@sce.carleton.ca

This work is supported in part by a Discovery Grant (DG) of the Natural Sciences and Engineering Research Council (NSERC) of Canada, in part by Huawei Canada Co., Ltd., and in part by the Ontario Ministry of Economic Development and Innovations ORF-RE (Ontario Research Fund–Research Excellence) program.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

joint optimization of various network aspects were made for relatively small networks, e.g., in [13]–[15] for joint scheduling and power allocation and in [16], [17] for joint scheduling, power allocation and routing. The common assumption in these attempts is that each subcarrier is used only once across the network. This assumption results in interference-free communication and facilitates implementation. However, it deprives the network from proper exploitation of its available resources. A better approach is to consider the possibility of reusing each subcarrier by multiple links over different time intervals. In this case, managing the interference resulting from subcarrier reuse can pose significant difficulty in designing the wireless network. An instance of managing this interference in single carrier systems was considered in [18]. Another instance in which interference is managed in relatively small multi-carrier systems appears in [19]. In that work, routes, powers and subcarrier schedules were jointly optimized by using an iterative Geometric Programming (GP) approximation of the original non-convex optimization problem. Simulation results reported in [19] suggest that subcarrier reuse enables better exploitation of resources and superior performance gains. However, these gains come at the expense of complexity. In particular, allowing each subcarrier to be used by all links results in a design complexity that grows exponentially with the size of the network. This renders the design approach developed in [19] overly complicated for usage in D2D networks. Hence for such networks, it is desirable to develop joint design approaches that provide close to optimal performance with a reasonable computational cost. This is the main focus of the first part of this paper.

In the second part, we focus mainly on resource allocation in large networks, i.e., networks with 100+ nodes. In such cases, joint optimization of routing, scheduling and power allocation across the whole network is computationally prohibitive. One approach to mitigate this difficulty is to cluster nodes into smaller groups such that the resource allocation problem is decomposed into smaller subproblems. This approach is widely used in wireless sensor networks, see e.g. [20], [21]. However in such networks, elaborate computations cannot be performed without heavily infringing on the typicall small battery-life of the sensors. Hence, in that work the resource allocation in each cluster is fairly simple and does not involve joint optimization. Another example for clustering is the study done for femtocells in [22]. In that paper, a semi-definite programming was used to cluster nodes and then an exhaustive search was used to find the best subcarrier scheduling and power allocation combination in each cluster. While the overall network setup resembles the one under consideration herein, the work in [22] did not consider routing and subcarrier sharing among users, which is the focus of the second part of this paper.

We consider an OFDMA-based D2D communication network in which the nodes are capable of sending, receiving and relaying data to other nodes [3]. Nodes acting as multihop relays operate in the half-duplex mode, i.e., a node cannot send and receive on the same subcarrier simultaneously. Nodes in the system are assumed to be connected to the BS through a control channel and the BS has access to the channel state information (CSI) of the nodes. Each subcarrier can be reused over multiple links. However, to simplify the design, at most one interferer is allowed within a geographic proximity at any time instant. This assumption is based on the fact that, in dense areas, reusing one subcarrier on more than two links results in severe interference and, subsequently, deteriorated performance. In addition to determining the power allocated for each transmission, resource-efficient communication between source-destination pairs in D2D networks requires judicious choice of the relaying nodes, the data routes, the subcarrier schedules and the fraction of time during which a subcarrier is assigned to a particular link. The problem of determining such decisions is NP-hard [12], and hence finding the optimal decisions is computationally infeasible for even small-to-moderate size networks.

The goal of this paper is to develop a joint optimization framework and a computationally efficient technique for designing wireless D2D communication networks with potentially tens of thousands of nodes. The optimization problem considered herein resembles the joint routing, scheduling and power allocation (JRSPA) considered in [19], but with a significantly larger number of nodes; the networks considered in [19] have tens of nodes, whereas the networks considered herein have thousands of nodes. This large number of nodes required a fundamentally different approach in solving the JRSPA optimization problem. In particular, whereas the solution of the JRSPA problem in [19] relied on GP and monomial approximations, which resulted in high complexity and slow convergence, the solution proposed herein relies on decomposing the JRSPA problem into two efficiently-solvable sub-problems, one for scheduling and routing and the other for power allocation. These sub-problems are solved in a two-stage iterative fashion, whereby the output of one sub-problem is used to obtain an initial point for the other sub-problem in the subsequent iteration. In the first stage, the power allocations are set to some fixed values and this causes the joint optimization of subcarrier schedules and routes to assume the form of an efficiently solvable linear program (LP). In the second stage, the output subcarrier schedules from the first stage are fixed and the power allocations and routes are optimized jointly. Unfortunately, the power allocation problem is non-convex. To overcome this difficulty we develop a novel iterative technique that we refer to as 'constraint-splitting'. This technique exhibits fast convergence and, in many cases, yields close to optimal power allocations within a small number of iterations. The philosophy of this technique is to split a particular constraint into two parts, each of which can be cast in a convex form. In particular, we observe that by fixing the right-hand-side (RHS) of that constraint and defining an appropriate lower bound, the problem can be cast as a GP, which can be readily converted into a convex optimization problem. We also observe that fixing the left-hand-side (LHS) of that constraint makes the problem convex and hence, efficiently solvable. We perform inner iterations over the fixed values on both sides of the constraint until convergence and we use the output in a steepest-descent outer iteration to update the subcarrier schedules in the first stage. Outer iterations continue until convergence. The two-stage algorithm exhibits a much less computational cost, and numerical results suggest that its performance is significantly better than existing joint routing, scheduling and power allocation techniques.

Despite the efficacy of the aformentioned two-stage algorithm, it is only capable of designing networks with hundreds of nodes. This is a substantial improvement over existing algorithms [19] which can only be used to design networks with at most ten nodes. However, with emerging smart applications, future wireless networks are envisioned to have thousands of nodes, and even more powerful joint design techniques are required. One such technique is proposed herein. In this technique, we consider a two-tier framework wherein the network is partitioned into several sub-networks. Each sub-network is assumed to have only a few hundred nodes and a gateway node, which can be one of the BSs in the cellular network. This gateway acts as a controller and a data aggregator. In the lower tier of the proposed framework, the gateway uses the two-stage algorithm to perform the joint optimization for nodes communicating within its own sub-network. For nodes communicating across sub-networks, the gateway aggregates the data that flows into and out of its sub-network and also relays data between source and destination gateways. The joint optimization of the network composed of gateways constitute the upper tier of the proposed framework. This optimization is performed by a global controller which uses the two-stage algorithm but with the gateways that control the sub-networks. Numerical examples confirm the superiority of this framework over the currently available techniques for designing large networks. In comparison with relevant work in the literature including our previous work in [17], [19], the main contributions of this work can be summarized as follows.

- We provide an iterative two-stage optimization approach for performing conjoint routing and resource allocation problems. This approach yields superior performance over currently available methods and with much less computational cost.
- We introduce a novel 'constraint-splitting' approach for the resource allocation problem in which a set of non-convex constraints is split into two sets of convex ones. We develop an efficient technique for iterating between formulations corresponding to the two sets of constraints.
- We develop a two-tier architecture, whereby the network is considered as a set of distinct clusters, each of which with a data aggregator that acts as a virtual source and/or destination for nodes in other clusters.
- Using the new two-stage algorithm with the constraint-splitting approach in the two-tier architecture enables efficient and effective design of networks with 1000+ nodes.

The paper is organized as follows. The system model and problem formulation are described in Section II. In Section III, two sub-problems are discussed, one for subcarrier scheduling when power allocations are fixed and one for power allocation when subcarrier schedules are fixed. In Section IV, the technique developed in Section III-B is used to develop a novel approach, which generates an approximate solution to the original design problem. The computational complexity of the proposed techniques is analyzed in Section V. The two-tier framework for designing large networks is presented in Section VI. In Section VII, simulation results are provided, and Section VIII concludes the paper.

II. System Model and Problem Statement

Performance of wireless communication networks depends on the interplay between network functionalities including end-to-end rate selection, data routing, time and frequency scheduling and power allocation. A model representing these interrelations is presented next.

A. System Model

We consider a communication network of N nodes, labelled $n = 1, \dots, N$ and L directed links, labelled $\ell = 1, \dots, L$. The sets of nodes and links are represented by $\mathcal{N}$ and $\mathcal{L}$, respectively. Each node has one transmit and one receive antenna and is capable of sending, receiving and relaying data to other nodes in the network. Data is communicated across the network through potentially multi-hop routes. We identify the data flows by their destinations. Let $\mathcal{D} \triangleq \{1, \dots, D\}$ be the set of destination nodes, $\mathcal{D} \subseteq \mathcal{N}$. For destination $d \in \mathcal{D}$, we use $s_n^{(d)}$ to denote the nonnegative end-to-end rate from node $n \in \mathcal{N}$ to destination $d \in \mathcal{D}$. Nodes are assumed to have finite power budget, $P_n,\ n = 1, \dots, N$, and infinite buffering capacity. We model the topology of this network by a directed graph in which nodes and links are represented by vertices and directed edges, respectively. We define $\mathcal{L}_+(n)$ and $\mathcal{L}_-(n)$ to be the set of links that are outgoing from and incoming to node $n \in \mathcal{N}$, respectively. The connection between nodes and links can be accounted for by the incidence matrix, $A \in \mathbb{R}^{N \times L}$, the entries of which are $a_{n\ell} = 1$ if $\ell \in \mathcal{L}_+(n)$, $a_{n\ell} = -1$ if $\ell \in \mathcal{L}_-(n)$ and zero otherwise. We consider the widely-used multicommodity flow model for the routing of data packets across the network, see, e.g., [11]. We assume that the data flows are lossless across links, and that the traffic flow can be split arbitrarily at nodes as long as the flow conservation law is satisfied at each node.

Using an OFDMA-based air-interface, the available frequency bandwidth, W, is divided into K narrowband subcarriers, each with a bandwidth of $W_0 = \frac{W}{K}$. The set of the K subcarriers is denoted by $\mathcal{K}$. Let $h_{\ell\ell'}^{(k)}$ represent the channel coefficient, which includes the path loss, shadowing and Rayleigh fading, on subcarrier k between the transmitter of link ℓ' and the receiver of link ℓ, $\ell, \ell' \in \mathcal{L}$, $k \in \mathcal{K}$. The network considered in this paper is quasi-static, which implies that $\{h_{\ell\ell'}^{(k)}\}$ remain constant over the signalling interval. This network can be represented by a graph in which each link has K distinct sublinks. We use $x_{\ell k}^{(d)}$ to denote the rate of data carried over subcarrier k of link ℓ and intended for destination node d, $\ell \in \mathcal{L}, k \in \mathcal{K}, d \in \mathcal{D}$. We also use $p_{\ell k}$ to denote the power used by the transmitter of link ℓ on subcarrier k.

To facilitate practical implementation, relaying nodes are assumed to operate in the half-duplex mode, whereby a node cannot simultaneously transmit and receive on the same subcarrier at the same time. Furthermore, it is assumed that a node cannot use the same subcarrier to broadcast different information to multiple nodes. However, it can do so either

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

on different subcarriers or at different time instances. As such, only one of the links in $\mathcal{L}_+(n)$, $n \in \mathcal{N}$, $k \in \mathcal{K}$, is potentially non-zero. Finally, a node can receive data from multiple nodes on the same subcarrier. In this case, it is possible for the node to use maximum likelihood or successive interference cancellation for joint detection, this approach is overly complicated and will not be considered in this paper. As such, we assume that receiving nodes use sequential detection while treating signals coming from other nodes as additive Gaussian noise. With this assumption, each link can be regarded as a single-user Gaussian channel with Shannon capacity $W_0 \log(1+\rho_{\ell k})$ where $\rho_{\ell k}$ is the received signal-to-interference-plus noise ratio (SINR) at the receiver of link ℓ on subcarrier k. This SINR is given by

$$\rho_{\ell k} = \frac{p_{\ell k}|h_{\ell\ell}^{(k)}|^2}{\sigma^2 + \sum_{\ell' \in \mathcal{L}\setminus\{\ell\}} p_{\ell' k}|h_{\ell\ell'}^{(k)}|^2}, \quad \ell \in \mathcal{L}, k \in \mathcal{K}, \tag{1}$$

where $\setminus$ represents the setminus operation and σ^2 represents the variance of the additive Gaussian noise at each receiving node. For simplicity, we will use $g_{\ell\ell'}^{(k)}$ to denote $\frac{|h_{\ell\ell'}^{(k)}|^2}{\sigma^2}$. A distinguishing feature of the interference expression in (1) is that it contains two parts: 1) the interference from other nodes communicating with the same receiver on subcarrier k; and 2) the interference from other nodes communicating with other receivers on subcarrier k.

B. Problem Formulation

We begin our analysis by considering the mathematical formulation for the joint routing, scheduling and power allocation problem developed in [19]. In this formulation, $s_n^{(d)}$, the data rate injected into the network at source node $n \in \mathcal{N}$ and intended for destination $d \in \mathcal{D}$, is assigned a prescribed nonnegative priority weight $w_n^{(d)}$, which can be changed over time to satisfy quality of service requirements. The collection of such weights are normalized so that $\frac{1}{ND}\sum_{n,d} w_n^{(d)} = 1$. The scheduling variables are characterized by the entries of the set $\Gamma = \{\gamma_{\ell_1,\ell_2,\ldots,\ell_m}^{(k)} | m = 1,\ldots,L, k = 1,\ldots,K\}$. These entries represent the fraction of time over which a particular subset of links utilizes the same subcarrier. For instance, $\gamma_{\ell_1,\ldots,\ell_m}^{(k)}$ denotes the fraction of the signalling interval during which links $\ell_1,\cdots,\ell_m \in \mathcal{L}$ are simultaneously 'active' on subcarrier $k \in \mathcal{K}$ and the remaining $L-m$ links in $\mathcal{L}$ are 'silent'on this subcarrier.

Using these notations, the JRSPA design problem can be cast in the following form:

$$\max_{\{s_n^{(d)}\},\{x_{\ell k}^{(d)}\},\{p_{\ell k}\},\{\gamma_{\ell\ell'}^{(k)}\}} \quad \sum_{d\in\mathcal{D}} \sum_{n\in\mathcal{N}\setminus\{d\}} w_n^{(d)} s_n^{(d)}, \tag{2a}$$

subject to $\Gamma \geq 0$, elementwise, (2b)

$$s_n^{(d)} \geq 0, \quad n \in \mathcal{N}\setminus d, d \in \mathcal{D}, \tag{2c}$$

$$x_{\ell k}^{(d)} \geq 0, \quad \ell \in \mathcal{L}, k \in \mathcal{K}, d \in \mathcal{D}, \tag{2d}$$

$$p_{\ell k} \geq 0, \quad \ell \in \mathcal{L}, k \in \mathcal{K}, \tag{2e}$$

$$\sum_{\ell\in\mathcal{L}} \sum_{k\in\mathcal{K}} a_{n\ell} x_{\ell k}^{(d)} = s_n^{(d)}, \quad n \in \mathcal{N}\setminus d, d \in \mathcal{D}, \tag{2f}$$

$$\sum_{m=1}^{L} \sum_{\ell_1\cdots\ell_m\in\mathcal{L}} \gamma_{\ell_1\ldots\ell_m}^{(k)} \leq 1, \quad k \in \mathcal{K}, \tag{2g}$$

$$a_{n\ell_1}^{+} a_{n\ell_2}^{-} \Big(\gamma_{\ell_1\ell_2}^{(k)} + \sum_{m=3}^{L} \sum_{\ell_3\cdots\ell_m\in\mathcal{L}} \gamma_{\ell_1\ldots\ell_m}^{(k)}\Big) = 0, \quad \ell_1 \in \mathcal{L}, \ell_2 \in \mathcal{L}\setminus\{\ell_1\}, k \in \mathcal{K}, \tag{2h}$$

$$a_{n\ell_1}^{+} a_{n\ell_2}^{+} \Big(\gamma_{\ell_1\ell_2}^{(k)} + \sum_{m=3}^{L} \sum_{\ell_3\cdots\ell_m\in\mathcal{L}} \gamma_{\ell_1\ldots\ell_m}^{(k)}\Big) = 0, \quad \ell_1 \in \mathcal{L}, \ell_2 \in \mathcal{L}\setminus\{\ell_1\}, k \in \mathcal{K}, \tag{2i}$$

$$\sum_{k\in\mathcal{K}} \sum_{\ell_1\in\mathcal{L}_+(n)} p_{\ell_1 k} \Big(\gamma_{\ell_1}^{(k)} + \sum_{m=2}^{L} \sum_{\ell_2\cdots\ell_m\in\mathcal{L}} \gamma_{\ell_1\ldots\ell_m}^{(k)}\Big) \leq P_n, \quad n \in \mathcal{N}, \tag{2j}$$

$$\sum_{d\in\mathcal{D}} x_{\ell_1 k}^{(d)} \leq \gamma_{\ell_1}^{(k)} \log_2(1 + p_{\ell_1 k} g_{\ell_1\ell_1}^{(k)}) + \sum_{m=2}^{L} \sum_{\ell_2\ldots\ell_m\in\mathcal{L}} \gamma_{\ell_1\ldots\ell_m}^{(k)} \log_2\left(1 + \frac{p_{\ell_1 k} g_{\ell_1\ell_1}^{(k)}}{1+\sum_{i=2}^{m} p_{\ell_i k} g_{\ell_1\ell_i}^{(k)}}\right), \quad \ell_1 \in \mathcal{L}, k \in \mathcal{K}. \tag{2k}$$

The significance of the constraints in (2) was described in detail in [19]; however, for completeness we now provide a brief explanation of each of these constraints. The non-negativity constraints in (2b) and (2e) are obvious. The constraint in (2f) ensures the flow conservation law at each node, i.e., incoming and outgoing flows of each node must be equal. The constraint in (2g) guarantees that the total usage of subcarrier k does not exceed the normalized signalling interval. The constraint in (2h) enforces the half-duplex operation of the system, whereby an incoming and outgoing links of node $n \in \mathcal{N}$ cannot be active at the same time on subcarrier $k \in \mathcal{K}$. In this equation $a_{n\ell_1}^{+}$ and $a_{n\ell_1}^{-}$ represent incoming and outgoing links of node n, respectively. In other words, $a_{n\ell_1}^{+} = 1$ when $a_{n\ell_1} = 1$ and $a_{n\ell_1}^{-} = 1$ when $a_{n\ell_1} = -1$. In this constraint all the schedules that correspond to simultaneous transmissions on consecutive links are set to zero. Similar argument holds for the broadcasting constraint in (2i), whereby of all the outgoing links of node $n \in \mathcal{N}$ only one can be active at any time instant on subcarrier $k \in \mathcal{K}$. The constraint in (2j) enforces the energy budget of a node and finally the constraint in (2k) guarantees that the communication rate of each link does not exceed its capacity. An observation that will prove pivotal in subsequent developments relies on the fact that each term of the capacity expression consists of two parts, the first part is related to the time during which only one transmission is scheduled on subcarrier k and the second part is related to the time during which more than one transmission are scheduled on that subcarrier.

The optimization problem in (2) is highly non-convex because of the constraints in (2j) and (2k) and hence, generally difficult to solve. An attempt to solve this problem was made in [19] which was based on GP. The iterative algorithm proposed therein, although finds an approximate solution with theoretically-proven polynomial complexity, its practical complexity is high and its convergence is generally

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

slow especially for medium-to-large networks. To circumvent these difficulties, in this paper we propose a low complexity algorithm that exhibits fast convergence even for large networks. Our approach is to decompose the optimization in (2) into two smaller sub-problems, one for scheduling and one for power allocation, with a partial coupling between them. Before we describe these sub-problems, in the next section we will introduce preliminary simplifications on the formulation in (2). In particular, we will show how the constraints in (2h) and (2i) can be eliminated from the problem. We also show that imposing constraints on the maximum number of simultaneous transmissions on a subcarrier can reduce the number of variables from being exponential in the number of nodes to being polynomial in it.

C. Preliminary Simplifications

One of the key constituents that contribute to the high complexity of solving (2) follows from the high cardinality of Γ. To see that, we note that the number of entries in Γ that are needed to characterize all possible transmission scheduling combinations on each subcarrier is given by

$$\sum_{i=1}^{L} \binom{L}{i} = (2^L - 1). \tag{3}$$

The i-th term in the summation corresponds to the number of ways a subcarrier can be allocated to i out of L possible links. Despite being comprehensive of all possibilities, this number results in overwhelming complexity for large networks.

To reduce the complexity of solving (2) for larger networks, we note that the half-duplex and broadcasting constraints depend only on the network graph, i.e., $\{a_{n\ell}\}$. Hence, ensuring that these constraints are satisfied can be effected prior to solving (2) . In fact, the constraints in (2h) and (2i) enforce some entries of Γ to be zero and hence, can be removed from the variable set. For instance, if ℓ_1 and ℓ_2 are incoming and outgoing links of node n, respectively, then the half-duplex constraint enforces all the entries of Γ involving ℓ_1 and ℓ_2, e.g., $\gamma_{\ell_1\ell_2}^{(k)}$ and $\gamma_{\ell_1\ell_2\ell_3}^{(k)}$ for all $\ell_3 \in \mathcal{L} \setminus \{\ell_1, \ell_2\}$, to be zero. Hence, these two constraints can be enforced by pruning the set Γ prior to solving (2). The pruning rule is as follows: for each ℓ_1 and $\ell_2 \in \mathcal{L}$, if either $a_{n\ell_1}^{+} a_{n\ell_2}^{+} = 0$ or $a_{n\ell_1}^{+} a_{n\ell_2}^{-} = 0$, the corresponding time-shares are removed from the set Γ.

To further reduce the number of effective entries in Γ, we note that in D2D communications, which lies at the focus of this paper, a subcarrier is less likely to be reused over a large number of links. This is due to the severe interference that such a reuse would result in. Hence, in D2D communications, it is expected that most of the gain of frequency-reuse can be mustered by considering only few simultaneous transmission over a subcarrier; increasing the reuse factor is likely to yield a marginal gain but with significantly higher complexity. For simplicity, we will consider the case in which a subcarrier can be used simultaneously by at most $I = 2$ links in some detail. However, the forthcoming analysis can be readily generalized to $I = 2, \cdots, L$.

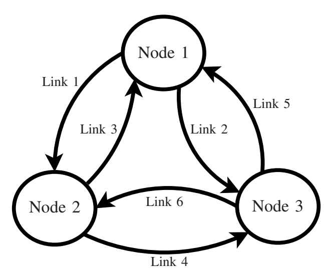
Using these simplifications, we now evaluate the number of entries in Γ. To do that, we note that when the reuse factor I=2, the entries of Γ can be arranged in the form of K, $L \times L$ matrices, $\{\Gamma^{(k)}\}_{k=1}^{K}$. The $\ell\ell'$-th entry of $\Gamma^{(k)}$ is given by $\gamma_{\ell\ell'}^{(k)}$, that is, this entry represents the fraction of the signalling interval over which subcarrier k is used on both links ℓ and ℓ'. Since, by definition, $\gamma_{\ell\ell'}^{(k)} = \gamma_{\ell'\ell}^{(k)}$, $\Gamma^{(k)}$ is symmetric. The diagonal entries of this matrix represent the fraction of time during which transmissions do not experience interference and its off-diagonal entries represent the fraction of time during which simultaneous transmissions interfere with each other.

For a fully connected graph, $L = N(N-1)$ and the cardinality of Γ can be reduced to

$$|\Gamma| = K\Bigg(N(N-1) + \binom{N(N-1)}{2} - N\binom{N-1}{2} - N\binom{N-1}{1}^2 + \frac{N(N-1)}{2}\Bigg) \tag{4}$$

$$= KN(N-1)\left(1 + \frac{(N-2)^2}{2}\right), \tag{5}$$

where the first and second terms in (4) represent the number of diagonal and (distinct) off-diagonal entries of Γ. The third term accounts for the variables that violate the half-duplex constraint, and the last two terms represent the number of variables that violate the broadcasting constraint. (The last term compensates for the variables that are counted twice in the preceding term.) Comparing (3) with (5), it can be seen that the preliminary simplifications proposed in this section reduces the cardinality of Γ from being exponential to being polynomial in N. An exemplary network of 3 nodes and its corresponding $\Gamma^{(k)}$ matrix is illustrated in Figure 2.



$$\Gamma^{(k)} = \begin{bmatrix} \gamma_{1,1}^{(k)} & \times & - & - & - & \gamma_{1,6}^{(k)} \\ \times & \gamma_{2,2}^{(k)} & - & \gamma_{2,4}^{(k)} & - & - \\ - & - & \gamma_{3,3}^{(k)} & \times & \gamma_{3,5}^{(k)} & - \\ - & \gamma_{2,4}^{(k)} & \times & \gamma_{4,4}^{(k)} & - & - \\ - & - & \gamma_{3,5}^{(k)} & - & \gamma_{5,5}^{(k)} & \times \\ \gamma_{1,6}^{(k)} & - & - & - & \times & \gamma_{6,6}^{(k)} \end{bmatrix},$$

$\times$ violates broadcast assumption,
$-$ violates half-duplex assumption.

Fig. 2: The scheduling matrix for a 3-node network with $L = 6$ links.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

III. Joint Design Sub-Problems: Scheduling and Power Allocation

In this section, we decouple the optimization problem in (2) into two parts, one for the scheduling when the power allocations are fixed and one for the power allocations when the schedules are fixed. After solving the two sub-problems, we will develop in Section IV an iterative technique to obtain a sub-optimal solution of the entire problem. For ease of exposition, the sub-problems will be described without invoking the simplifications in the previous section. However, in the numerical examples, these simplifications will be used to reduce the complexity of the design problem.

A. Scheduling With Fixed Power Allocations

In this section, we consider the problem of optimizing the subcarrier schedules that maximize a weighted-sum rate of the network when the power allocations are fixed. Let $\tilde{q}_{\ell k}^{(k)}$ denote the power allocations which are assumed to be fixed in this phase. Careful examination of the optimization problem in (2) reveals that with the power allocations fixed, this problem becomes an LP and hence, its global maximum can be found in polynomial time, cf. e.g., [16].

B. Power Allocation With Fixed Schedules

In this section, we consider a problem complementary to the one presented in the previous section, i.e., the problem of optimizing the power allocations that maximize a weighted-sum rate of the network when the subcarrier schedules are fixed. Let $\tilde{\Gamma}$ denote the schedules which are assumed to be fixed in this phase. The problem in (2) with fixed schedules reduces to a GP at high SINR regimes. In those regimes, the solution can be found optimally [23]. However, the problem with this approach is that the SINR are not known prior to performing power allocation, and, in general, this problem is non-convex and difficult to solve. In [23], an iterative technique based on monomial approximation was used to find a suboptimal solution. However, in [19] it was shown that the convergence of this technique is relatively slow, for all but the smallest of networks, which renders the approach in [23] impractical for medium-to-large networks. To tackle this problem, we develop a novel approach, which we refer to as 'constraint-splitting'. This approach will be shown in Section V-C to exhibit significantly faster convergence.

We begin the development of the proposed approach by expressing the capacity constraint in (2k) in a format that is more amenable to prospective optimization. We will focus on the case of having at most two interferers, i.e., $I = 2$. However, the forthcoming formulations can be readily extended to cover cases with $I > 2$, but unfortunately, not without compromising clarity of exposition.

$$\sum_{d\in\mathcal{D}} x_{\ell k}^{(d)} \leq \gamma_{\ell\ell}^{(k)} \log(1+p_{\ell k} g_{\ell\ell}^{(k)}) + \sum_{\ell'\in\mathcal{L}\setminus\{\ell\}} \gamma_{\ell\ell'}^{(k)} \log\left(1+\frac{p_{\ell k} g_{\ell\ell}^{(k)}}{p_{\ell' k} g_{\ell\ell'}^{(k)}}\right). \tag{6}$$

The next step is to rewrite the capacity constraint in (6) in a form that facilitates the optimization of the power allocations. This constraint can be written as:

$$\sum_{d\in\mathcal{D}} x_{\ell k}^{(d)} + \underbrace{\sum_{\ell'\in\mathcal{L}\setminus\{\ell\}} \gamma_{\ell\ell'}^{(k)} \log\left(1+p_{\ell' k} g_{\ell\ell'}^{(k)}\right)}_{\text{Interference part}} \leq \underbrace{\gamma_{\ell\ell}^{(k)} \log(1+p_{\ell k} g_{\ell\ell}^{(k)}) + \sum_{\ell'\in\mathcal{L}} \gamma_{\ell\ell'}^{(k)} \log\left(1+p_{\ell' k} g_{\ell\ell'}^{(k)}\right)}_{\text{Noisy-signal part}}. \tag{7}$$

We will refer to the second summation on the RHS of (7) as *the interference part* because it contains the interference terms only and the summation on the LHS of (7) as *the noisy-signal part* because it contains both signal and interference terms.

Looking back into the optimization in (2) with fixed schedules and (2k) being replaced with (7), we make the following observation, which will later help us in proposing a fast-converging technique to solve the power allocation problem. The first observation is that, if we fix the noisy-signal part, the optimization problem in (2) can be cast in a GP form that can be easily converted into a convex problem [24]. The second observation is that, if we fix the interference part, the optimization problem in (2) reduces to a convex problem that can be efficiently solved using interior point methods [25]. Taking advantage of these observations, in the next two sections we will explain each of the aforementioned problems and then we will develop an iterative technique that exhibits fast convergence to a power allocation solution.

1) Interference Sub-Problem: In this section, we consider the problem in (2) when (2k) is replaced with (7) and the schedules are fixed. We will denote these schedules by $\{\tilde{\gamma}_{\ell\ell'}^{(k)}\}$ which are, in fact, the entries of $\tilde{\Gamma}$. Suppose that the RHS of (7) is fixed to some initial power allocation $\{p_{\ell k}^{(0)}\}$ and let us introduce a parameter $\alpha \geq 1$ which we will use to control the search region for a proper power allocation around $\{p_{\ell k}^{(0)}\}$. After fixing the RHS of (7) and introducing the parameter α, the joint design problem can be written in the following form:

$$\max_{\{s_n^{(d)}\},\{x_{\ell k}^{(d)}\},\{p_{\ell k}\}} \sum_{d\in\mathcal{D}} \sum_{n\in\mathcal{N}\setminus\{d\}} w_n^{(d)} s_n^{(d)}, \tag{8a}$$

$$\text{subject to } s_n^{(d)} \geq 0, \quad n\in\mathcal{N}\setminus d, d\in\mathcal{D}, \tag{8b}$$

$$x_{\ell k}^{(d)} \geq 0, \quad \ell\in\mathcal{L}, k\in\mathcal{K}, d\in\mathcal{D}, \tag{8c}$$

$$p_{\ell k} \geq 0, \quad \ell\in\mathcal{L}, k\in\mathcal{K}, \tag{8d}$$

$$\sum_{\ell\in\mathcal{L}} \sum_{k\in\mathcal{K}} a_{n\ell} x_{\ell k}^{(d)} = s_n^{(d)}, \quad n\in\mathcal{N}\setminus d, d\in\mathcal{D}, \tag{8e}$$

$$\sum_{k\in\mathcal{K}} \sum_{\ell\in\mathcal{O}(n)} p_{\ell k} \sum_{\ell'\in\mathcal{L}} \tilde{\gamma}_{\ell\ell'}^{(k)} \leq P_n, \quad n\in\mathcal{N}, \tag{8f}$$

$$\sum_{d\in\mathcal{D}} x_{\ell k}^{(d)} + \sum_{\ell'\in\mathcal{L}\setminus\{\ell\}} \tilde{\gamma}_{\ell\ell'}^{(k)} \log\left(1+p_{\ell' k} g_{\ell\ell'}^{(k)}\right) \leq \alpha S_{\ell k}, \quad \ell\in\mathcal{L}, k\in\mathcal{K}, \tag{8g}$$

where $S_{\ell k} = \tilde{\gamma}_{\ell}^{(k)} \log(1+p_{\ell k}^{(0)} g_{\ell\ell}^{(k)}) + \sum_{\ell'\in\mathcal{L}} \tilde{\gamma}_{\ell\ell'}^{(k)} \log\big(1+p_{\ell' k}^{(0)} g_{\ell\ell'}^{(k)}\big)$ is the fixed noisy-signal part.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

While the optimization problem in (2) is NP-hard, the one in (8) is in the form of an efficiently solvable GP. There is caveat though: solving (8) will result in maximizing the first summation on the LHS of (8g) at the expense of the link power $\{p_{\ell' k}\}$, thereby pushing them towards zero. This undesirable situation can be prevented by introducing the following lower bounds on the powers:

$$B_{\ell k} \leq \sum_{\ell' \in \mathcal{L} \setminus \{\ell\}} \tilde{\gamma}_{\ell\ell'}^{(k)} \log\big(p_{\ell' k} g_{\ell\ell'}^{(k)}\big), \qquad \ell \in \mathcal{L}, k \in \mathcal{K}, \quad (9)$$

where $B_{\ell k} = \sum_{\ell' \in \mathcal{L} \setminus \{\ell\}} \tilde{\gamma}_{\ell\ell'}^{(k)} \log\big(p_{\ell' k} g_{\ell\ell'}^{(k)}\big), \ell \in \mathcal{L}, k \in \mathcal{K}$ are constants obtained from the link powers generated in the preceding iteration. We note that one of the advantages of the bound in (9) is that it complies with the GP framework and hence can be readily incorporated in the framework in (8). The resulting optimization problem in this case can now be expressed as:

$$\max_{\{s_n^{(d)}\},\{x_{\ell k}^{(d)}\},\{p_{\ell k}\}} \quad \sum_{d \in \mathcal{D}} \sum_{n \in \mathcal{N} \setminus \{d\}} w_n^{(d)} s_n^{(d)}, \quad (10a)$$

$$\text{subject to} \quad \text{Constraints (8b)–(8g) and (9).} \quad (10b)$$

The sub-problem in (10) can be cast in a convex form by first using a logarithmic change of variables to write (10) in a form that conforms to the GP framework [25], which, using a standard exponential transformation, can be converted into a convex optimization problem [24].

2) Noisy-Signal Sub-Problem: In this section, we consider a case complementary to the one considered in the previous section, i.e., the case when interference part on LHS of (7) is fixed and the noisy-signal part is optimized. Again, we assume that an initial power allocation, $\{p_{\ell k}^{(0)}\}$, is given. Analogous to the discussion in the previous section, for the noisy-signal sub-problem we introduce a parameter $\beta \leq 1$ to control the search region for the power allocation around the given initial point. Given β, the noisy-signal sub-problem can be expresses as

$$\max_{\{s_n^{(d)}\},\{x_{\ell k}^{(d)}\},\{p_{\ell k}\}} \sum_{d \in \mathcal{D}} \sum_{n \in \mathcal{N} \setminus \{d\}} w_n^{(d)} s_n^{(d)}, \quad (11a)$$

$$\text{subject to } s_n^{(d)} \geq 0, \qquad n \in \mathcal{N} \setminus d, d \in \mathcal{D}, \quad (11b)$$

$$x_{\ell k}^{(d)} \geq 0, \qquad \ell \in \mathcal{L}, k \in \mathcal{K}, d \in \mathcal{D}, \quad (11c)$$

$$p_{\ell k} \geq 0, \qquad \ell \in \mathcal{L},\ k \in \mathcal{K}, \quad (11d)$$

$$\sum_{\ell \in \mathcal{L}} \sum_{k \in \mathcal{K}} a_{n\ell} x_{\ell k}^{(d)} = s_n^{(d)},\ n \in \mathcal{N} \setminus d, d \in \mathcal{D}, \quad (11e)$$

$$\sum_{k \in \mathcal{K}} \sum_{\ell \in \mathcal{O}(n)} p_{\ell k} \sum_{\ell' \in \mathcal{L}} \tilde{\gamma}_{\ell\ell'}^{(k)} \leq P_n, \qquad n \in \mathcal{N}, \quad (11f)$$

$$\sum_{d \in \mathcal{D}} x_{\ell k}^{(d)} + \beta I_{\ell k} \leq \tilde{\gamma}_{\ell}^{(k)} \log(1 + p_{\ell k} g_{\ell\ell}^{(k)})$$
$$+ \sum_{\ell' \in \mathcal{L}} \tilde{\gamma}_{\ell\ell'}^{(k)} \log\left(1 + p_{\ell' k} g_{\ell\ell'}^{(k)}\right),\ \ell \in \mathcal{L}, k \in \mathcal{K}, \quad (11g)$$

where $I_{\ell k} = \sum_{\ell' \in \mathcal{L} \setminus \{\ell\}} \tilde{\gamma}_{\ell\ell'}^{(k)} \log\left(1 + p_{\ell' k}^{(0)} g_{\ell\ell'}^{(k)}\right)$ is the fixed interference part. This problem is convex and can hence be readily solved with highly efficient interior-point method solvers.

To summarize, we emphasize that the way in which the constraint in (7) is split resulted in two convex optimization problems, the GP-compatible one in (10) and the one in (11). This is the key that will ensure efficient implementation of the iterative technique described next.

3) Iterative Solution for Power Allocation Sub-problem: We now develop an iterative technique that incorporates the convex problems described in Section III-B to solve the problem in (2) when the schedules are fixed.

Starting from a feasible initial power allocation, we first solve the interference sub-problem in (10) for a value of $\alpha > 1$. The solution is then used as an initial point for the signal sub-problem in (11) with a value of $\beta < 1$. The output of this sub-problem is then used as an initial point for the subsequent iteration. For this technique to converge, the feasible region must be expanded less at each iteration in order for the outputs of both the noisy-signal and the interference sub-problems to converge. This goal can be achieved by adjusting the parameters α and β at each iteration. In particular, for convergence, the value of α and β at the i-th iteration, α_i and β_i, respectively, must satisfy $\alpha_i \leq \alpha_{i-1}$ and $\beta_i \geq \beta_{i-1}$. At convergence, we must have $\alpha^* = \beta^* = 1$. It is worth noting that the step size for adjusting α and β must be not too small, to avoid slow convergence, and not too large to avoid crossing over of the powers generated by the two sub-problems. This algorithm is summarized in Algorithm 1.

Algorithm 1: Inner iteration: Constraint-splitting approach

Data: Subcarrier schedules, CSI, weights, initial power allocation
Result: data rates, power allocations
Initialization: set α and β;
while $\alpha \neq \beta$ **do**
    Solve the interference sub-problem (GP) in (10);
    Set the solution as the initial power allocation;
    Solve the noisy-signal sub-problem (Convex) in (11);
    Update the parameters α and β;
end

IV. Approximated Solution for Joint Design Problem

In the previous section, we considered the joint optimization problem when either the schedules or the powers are fixed. Using the techniques developed in Sections III-A and III-B, in this section, we provide an efficient technique for generating 'good' solutions of the joint optimization problem in (2) in its entirety. Our approach is composed of two stages, one for solving the joint optimization problem with fixed powers, and one for solving it with fixed schedules. Iterating between these two stages yields an approximate solution for the joint design problem in (2). It is worth noting that, in contrast to the GP-based approach in [19], this algorithm has much less computational complexity as it needs fewer iterations for convergence, cf. Section V below.

In the algorithm presented herein, we begin from a feasible initial point, for instance, equal distribution of the power budget among outgoing links, i.e., $p_{\ell k}^{(0)} = \frac{P_n}{K|\mathcal{O}(n)|}$, $n = 1, \cdots, N$. In the first stage, we fix the power allocation in (2) to $\{p_{\ell k}^{(0)}\}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

We then solve the resulting LP to find the optimal schedules, $\{\widetilde{\gamma}_{\ell\ell'}^{(k)}\}$, corresponding to the initial power allocation. In the second stage, we fix the schedules in (2) to $\{\widetilde{\gamma}_{\ell\ell'}^{(k)}\}$. We then use Algorithm 1 to find the corresponding power allocations $\widetilde{p}_{\ell k}$. These power allocations can be fed back into the first stage to solve the problem iteratively. However, our numerical results suggest that, in its current form, this outer iteration provides negligible performance gain. The reason is that schedules and powers are tightly coupled, i.e., if one of the powers is zero, the corresponding schedules are also zero and vice versa. To circumvent this difficulty, in the proposed algorithm we modify the power allocations in the outer iteration in order to enable further exploration of the feasible region. To do that, we use the gradient method [25] to find the gradient ascent direction of the problem in (2). In particular, we use the log-barrier method [25] to incorporate the inequality constraints in (2) in the objective. In this method, the problem in (2) is written in the following form:

$$\begin{aligned} \max \sum_{n,d} w_n^{(d)} s_n^{(d)} &+ \frac{1}{t}\Big(\sum_{n,d}\log(s_n^{(d)}) + \sum_{\ell,k,d}\log(x_{\ell k}^{(d)}) \qquad (12) \\ &+ \sum_{\ell,k}\log(\gamma_{\ell\ell'}^{(k)}) + \sum_{\ell,k}\log(p_{\ell k})\Big) + \frac{1}{t}\sum_{\ell,k}\log\Big(\psi(\ell,k)\Big) \\ &+ \frac{1}{t}\sum_{k}\log(1-\sum_{\ell\,\ell'\in\mathcal{L}}\gamma_{\ell\ell'}^{(k)}) + \frac{1}{t}\sum_{n}\log\big(\phi(n)\big), \\ \text{subject to } &\sum_{\ell,k} a_{n\ell}x_{\ell k}^{(d)} = s_n^{(d)}, \quad n\in\mathcal{N}\setminus d, d\in\mathcal{D}, \end{aligned}$$

where $\phi(n) \triangleq P_n - \sum_{k,\ell\in\mathcal{O}(n)} p_{\ell k}\sum_{\ell'}\gamma_{\ell\ell'}^{(k)}$, $\psi(\ell,k) \triangleq \gamma_{\ell\ell}^{(k)}\log(1+p_{\ell k}g_{\ell\ell}^{(k)}) + \sum_{\ell'}\gamma_{\ell\ell'}^{(k)}\log\big(1+\frac{p_{\ell k}g_{\ell\ell}^{(k)}}{1+p_{\ell' k}g_{\ell\ell'}^{(k)}}\big) - \sum_d x_{\ell k}^{(d)}$ represent the gap between the RHS and LHS of the constraints in (2j) and (2k), respectively, and t represents the log-barrier parameter. The gradient of the objective in (12) with respect to $p_{\ell k}$ can be readily shown to be given by

$$\begin{aligned}\nabla_{p_{\ell k}} = \frac{1}{t}\Bigg(&\frac{1}{p_{\ell k}} - \frac{\sum_{\ell\in\mathcal{L}_+(n),k,\ell'}\gamma_{\ell\ell'}}{\phi(n)} \\ &+ \frac{\frac{\gamma_{\ell\ell}^{(k)}g_{\ell\ell}^{(k)}}{1+p_{\ell k}g_{\ell\ell}^{(k)}} + \sum_{\ell'}\frac{\gamma_{\ell\ell'}^{(k)}g_{\ell\ell}^{(k)}\left(1+g_{\ell\ell'}^{(k)}(p_{\ell' k}-p_{\ell k})\right)}{(1+p_{\ell' k}g_{\ell\ell'}^{(k)})^2}}{\psi(\ell,k)}\Bigg). \qquad (13)\end{aligned}$$

We set $t = 1$ in the first iteration and increase it in subsequent ones. We use the gradient ascent direction in (13) to update the output of the second stage and feed it back into the first stage in the following iteration. In particular, we use the following update rule:

$$\tilde{p}_{\ell k}^{(j+1)} = \tilde{p}_{\ell k}^{(j)} + \mu_j \nabla_{\tilde{p}_{\ell k}}, \qquad \ell\in\mathcal{L}, k\in\mathcal{K}, \qquad (14)$$

where j is the index of outer iterations and μ_j is a step size. Iterations continue until a stopping criterion is satisfied, e.g., no significant improvement in the objective is observed. For guaranteed convergence [25], $\{\mu_j\}$ are chosen to form a monotonically decreasing sequence that satisfies $\sum_j \gamma_j = \infty$. Details are summarized in Algorithm 2 and illustrated in Figure 3.

Algorithm 2: Outer iteration: Approximated solution for JRSPA problem in (2)

Data: CSI, weights
Result: data rates, subcarrier schedules, power allocations
Initialization: set $\{p_{\ell k}^{(0)}\}$ as the equal power assignment;
while $\|\nabla\| > \epsilon$ **do**
  Stage 1: solve (2) with fixed powers to find subcarrier schedules (LP);
  Stage 2: run Algorithm 1 to obtain power allocation, $\{\widetilde{p}_{\ell k}\}$;
  Update: use (14) to update the obtained powers;
end

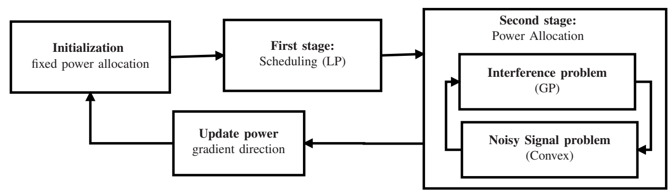


Fig. 3: Block diagram of Algorithm 2.

It will be shown in Section VII that Algorithm 2 yields solutions that perform significantly better than those yielded by a fixed power allocation approaches. This algorithm will also be compared to the GP-based approach in [19]. Furthermore, it will be shown that this algorithm tends to yield a better performance with significantly less computational cost.

In the next section, we will provide bounds on the computational complexity of the proposed techniques. In particular, we will show that each stage of the algorithm has a polynomial complexity and hence, the proposed algorithm for obtaining an approximate solution to the joint optimization problem in (2) also has a polynomial complexity.

V. Computational Complexity

The approach proposed in the previous section is based on iterating between two stages. In the first stage, we seek the optimal schedules for a given power allocation, whereas in the second stage, we find a suboptimal power allocation for a given schedules. The complexity of each stage is discussed next.

A. Computational Complexity of the First Stage

In the first stage of the approach proposed in the previous section, the power allocations are fixed. In this case, the problem in (2) reduces to an LP where the optimal solution, i.e., optimal schedules, could be found efficiently using IPM-based solvers. The number of Newton iterations required by such solvers can be shown to be proportional to $\sqrt{m}$, where m is the number of inequality constraints [25]. For the LP problem, we have $m = LK(D+1) + D(N-1) + N + K + KN(N-1)\Big(1+\frac{(N-2)^2}{2}\Big)$. In addition, each Newton step is known to have a cubic complexity [26]. Hence, in the worst case scenario when the network is fully connected, i.e., $L = N(N-1)$ and all the nodes are destination nodes,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

i.e., $D = N$, the computational complexity of solving the LP problem is $\mathcal{O}(\frac{1}{2}K^{3.5}N^{14})$.

B. Computational Complexity of the Second Stage

In the second stage, sub-optimal power allocations for given schedules are obtained by solving a sequence of convex problems. The complexity of each problem is discussed next.

1) Computational Complexity of the Interference Sub-Problem: The interference sub-problem discussed in Section III-B1 yields a GP which can be readily converted into a convex problem using the exponential change of variables [24]. The computational complexity of solving such problems were studied in [17], [19] by bounding each monomial term in the GP with a new variable that serves as an upper bound. Using this method, it can be shown that the complexity of solving the GP problem in worst case scenario is $\mathcal{O}(4K^{3.5}N^{14})$.

2) Computational Complexity of the Noisy-Signal Sub-Problem: The noisy-signal sub-problem discussed in Section III-B2 yields a convex optimization problem. Using a discussion analogous to the one in Section V-A, it can be shown that the number of inequality constraints is $m = LK(D + 2) + D(N - 1) + N$. Hence, the complexity of solving this problem in the worst case scenario using IPM-based solvers is $\mathcal{O}(K^{3.5}N^{10.5})$.

C. Computational Complexity of the Two-Stage Approach

We begin by recalling that the parameters α and β control the search region for a power allocation solution. Let ϵ be the step size with which the parameter α shrinks at each inner iteration of the algorithm and let $\beta = \frac{1}{\alpha}$. Since at convergence, we must have $\alpha = 1$, the number of iterations required for convergence is $\frac{\alpha}{\epsilon}$. Using the two-stage approach presented in Section IV and the complexity discussions in Sections V-B1 and V-B2, it can be seen that the complexity of each outer iteration of the proposed approach is bounded by

$$\mathcal{O}\Big(K^{3.5}N^{10.5}\big(\frac{\alpha}{\epsilon}(4N+1)\big) + \frac{1}{2}\Big). \tag{15}$$

We conclude this section by noting that, when the reuse factor, I, is restricted to be small, both the method proposed herein and the one proposed in [19] have polynomial complexity. However, the main difference between these methods is that the typical number of iterations required for the method proposed in [19] to converge is much larger than its counterpart for the method proposed herein. For instance, using the method proposed in [19] in a network with $N = 4$ nodes required 180 iterations to converge. This is in contrast with the method proposed herein, in which we are able to find a sub-optimal solution within less than 10 iterations.

VI. Joint Design in Large Networks

The algorithm proposed in Section IV enables us to jointly design the data routes, subcarrier schedules and power allocations in networks of hundreds nodes. Although this is a significant improvement compared to the algorithm in [19] where the joint design is applicable to networks of up to ten nodes, in practice, the size of data networks might be much larger and it is desirable to solve the joint design problem for such networks.

In this section we consider large networks with thousands of nodes. From the complexity analysis presented in Section V-C, it can be seen that, despite the efficacy of the joint design algorithm presented in Section IV, using this algorithm to design a network with thousands of nodes in one shot is computationally prohibitive. To circumvent this difficulty, we propose a two-tier communication framework. We begin by assuming that the network is composed of several disjoint clusters, which are not necessarily far from each other in a geographical sense. Each cluster has a local cluster controller (CC), which can be either an entity outside the network or one of the nodes within the cluster. Communications between CCs in the higher tier is controlled by a central entity which we refer to it as the global controller (GC). We assume that communication between CCs is performed over a set of frequencies than are distinct from the set used for communication between nodes in the network. Such a scenario arises naturally in heterogeneous networks with one macro and multiple femto BSs. In these networks, the femto BSs act as CCs that are responsible for accommodating communications within their cells and the macro BS acts as a GC that is responsible for communication between femto cells [27].

An exemplary network of four clusters is illustrated in Figure 4. In this figure, each cluster has multiple BSs but only one of them is designated as a CC (marked in red). A fraction of the nodes in this network wish to communicate with nodes within their cluster (marked in green) and another fraction of nodes wish to communicate with nodes outside their cluster (marked in blue). The frequencies used for communications between CCs (marked with dashed lines) are distinct from those used for communication inside clusters (marked with solid lines). The communication between CCs is controlled by the GC through a control channel (dotted lines).

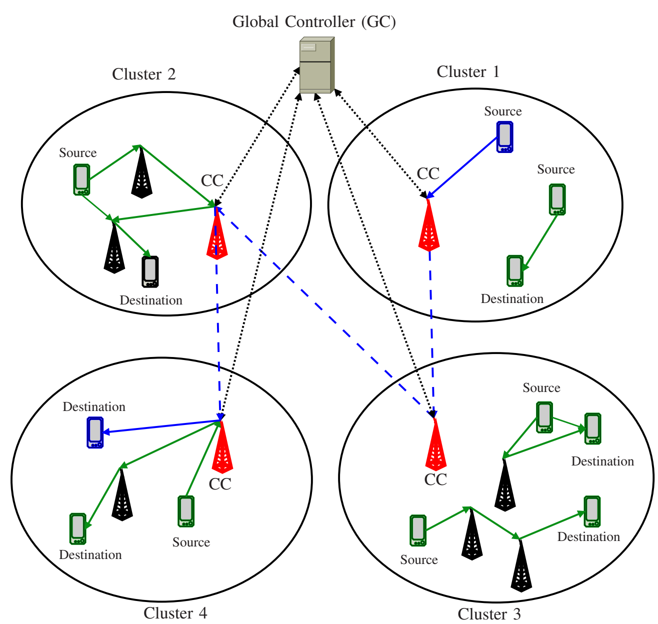


Fig. 4: An exemplary network of 4 clusters.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

A. Proposed Framework

We consider a network of M clusters, which form the set $\mathcal{M} \triangleq \{1, \ldots, M\}$. The sets of nodes and destinations in cluster $i \in \mathcal{M}$ are denoted by $\mathcal{N}_i$ and $\mathcal{D}_i$, respectively. At the beginning of each scheduling interval, source nodes announce their intended destinations to their respective CCs. Communication between nodes in the network falls in one of two categories:

- Intra-Cluster: This case arises when both the source and destination nodes are located within the same cluster. The CC of the cluster incorporates the parameters of these nodes in the optimization framework of Section IV. The schedules, routes and power allocations output of this optimization are then passed by the CC to the nodes within its cluster.
- Inter-Cluster: This case arises when the source and destination nodes are located in different clusters. The CCs of the source and destination clusters are responsible for pulling data from the source node and pushing it to the destination node, respectively. Communication between the CCs of the source and destination clusters is handled by the GC. In particular, the GC uses the framework of Section IV to find the optimized schedules, routes and power allocations for the communication over the network of CCs.

We now use these two categories to describe three phases of the proposed framework.

1) First Phase (Intra-Cluster Design): In this phase, the CCs perform three tasks: first, they perform the joint design for the nodes that lie within their clusters; second, they act as virtual destination nodes for any source whose actual intended destination lies outside the cluster; and third, they act as virtual source nodes for any destination within the cluster but whose actual source lies outside the cluster. In the latter two tasks, the CCs act as gateways for their respective clusters. Each CC uses the algorithm in Section IV to jointly design the routes, schedules and power allocations for the network formed by the nodes within its cluster, including the CC itself.

To characterize the role of CCs as gateways for their clusters, we begin by noting that, similar to other nodes in the network, a CC can act both as a source and a destination at the same time. Now, let $\mathcal{D}_i^{in}$ be the set of destination nodes whose source nodes lie inside cluster $i, i \in \mathcal{M}$. Also, let $\mathcal{N}_i^{out}$ be the set of source nodes whose destination lies outside cluster i and $\mathcal{D}_i^{out}$ be the set of nodes whose source nodes lie outside cluster i. These three sets, i.e., $\mathcal{D}_i^{in}$, $\mathcal{N}_i^{out}$ and $\mathcal{D}_i^{out}$, are illustrated in Figure 5. For $n \in \mathcal{N}_i^{out}$, CC_i acts as the destination node and for $d \in \mathcal{D}_i^{out}$, it acts as the source node. Hence, as before, we will use $s_n^{(CC_i)}$ to denote the data rate from the source node $n \in \mathcal{N}_i^{out}$ to its *virtual* destination, CC_i and $s_{CC_i}^{(d)}$ to denote the data rate from the *virtual* source node, CC_i to its corresponding destination, $d \in \mathcal{D}_i^{out}$. Now, CC_i performs the optimization in (2) to find the routes, schedules and power allocations. In particular, CC_i solves the optimization in (2) with the modified objective $\sum_{d\in\mathcal{D}_i^{in}} \sum_{n\in\mathcal{N}_i} s_n^d + \sum_{n\in\mathcal{N}_i^{out}} s_n^{CC_i} + \sum_{d\in\mathcal{D}_i^{out}} s_{CC_i}^{(d)}$. The first part of this objective accounts for the communications whose source and destination are located inside the cluster i, the second part accounts for the communications whose source and destination are inside and outside of cluster i, respectively, and the last part accounts for the communications whose source and destination are outside and inside cluster i, respectively. The constraints in (2) can be readily modified to define the feasible set of the new variables. To ensure that CC_i acts as a virtual source and destination node, the variables $\{s_n^d | n \in \mathcal{N}, d \in \mathcal{D}\}$ in (2) are replaced with $\{s_n^d | n \in \mathcal{N}_i \cup \{CC_i\}, d \in \mathcal{D}_i \cup \{CC_i\}\}$.

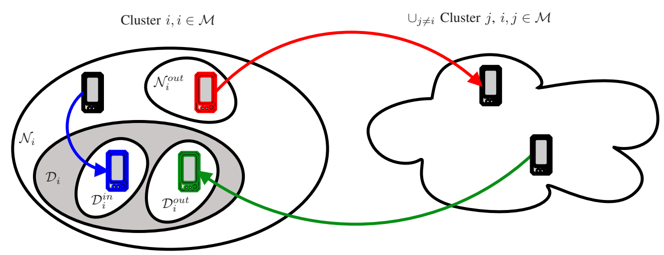


Fig. 5: Three sets of nodes in Cluster i: 1) destination nodes whose source is inside the cluster, $\mathcal{D}_i^{in}$ in blue, 2) destination nodes whose source is outside the cluster, $\mathcal{D}_i^{out}$ in red, and 3) source nodes whose destination is outside the cluster, $\mathcal{N}_i^{out}$ in green. Note that these sets are not necessarily disjoint.

2) Second Phase (Inter-Cluster Design): Here we consider the situation when the source and destination nodes lie in distinct clusters. This situation was handled in part in the previous phase. In particular, that phase is responsible for establishing communication between the source node and its CC, which acts as a virtual destination, and for establishing communication between the CC, which acts as a virtual source, and the destination node. Now, we consider the communication between CCs that act as virtual sources and the CCs that act as virtual destinations. This communication is coordinated by the GC.

To characterize the current phase, let $n \in \mathcal{N}_i$ be a source in cluster i and let $d \in \mathcal{D}_j$ be a destination node in cluster j, $i \neq j$, $i, j \in \mathcal{M}$. The goal in this phase is to establish communication between CC_i and CC_j. Noting that CC_i and CC_j serve as a proxy source and destination for the rate $s_n^{(d)}$, it can be seen that, with the proposed scheme, the data rate between the source $n \in \mathcal{N}_i$ and the destination $d \in \mathcal{D}_j$ cannot exceed $\min\{s_n^{(CC_i)}, s_{CC_j}^{(d)}\}$. Hence, we must have $s_{CC_i}^{(CC_j)} \leq \min\{s_n^{(CC_i)}, s_{CC_j}^{(d)}\}$, where $s_n^{(CC_i)}$ and $s_{CC_j}^{(d)}$ are obtained from solving the optimization in phase 1. Hence, to establish communication between CC_i and CC_j, the GC solves the following variation of (2) to obtain $\{s_{CC_i}^{(CC_j)}\}$ along with the respective schedules, routes and power allocations:

$$\max \sum_{i\in\mathcal{M}} \sum_{j\in\mathcal{M}\setminus\{i\}} s_{CC_i}^{(CC_j)}, \tag{16a}$$

subject to

$$\text{constraints in (2b)–(2k)}, \tag{16b}$$

$$s_{CC_i}^{(CC_j)} \leq \min\{s_n^{(CC_i)}, s_{CC_j}^{(d)}\},\ i \in \mathcal{M}, j \in \mathcal{M}\setminus\{i\}. \tag{16c}$$

We now make two remarks regarding the formulation in (16). First, we note that, in this formulation, we only considered

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

the total throughput, rather than the weighted sum of rates. This is because data flows passing through a CC may have different weights, and combining these weights for inter-cluster communications appears to be rather complicated. Second, we note that because $\min\{s_n^{(CC_i)}, s_{CC_j}^{(d)}\}$ is obtained from the solution of the first phase, the constraint in (16c) is convex and hence, incorporating it in the formulation does not reduce its solvability.

3) Third Phase (Update Intra-Cluster Design): After solving (16) in phase 2 for the rates $\{s_{CC_i}^{(CC_j)}\}$, these rates are communicated to the CCs and the network design process can be considered complete. However, we note that this design is amenable to further refinement. In particular, the rates obtained in phase 2, i.e., $\{s_{CC_i}^{(CC_j)}\}$, can be regarded as the end-to-end rates, and hence, there is no benefit in having either $s_n^{(CC_i)}$ or $s_{CC_j}^{(d)}$ exceed $s_{CC_i}^{(CC_j)}$. In other words, a refinement of the design can be obtained by re-solving (2), but with $s_n^{(CC_i)}$ and $s_{CC_j}^{(d)}$ bounded by $s_{CC_i}^{(CC_j)}$, i.e., with following constraints included in the formulation in (2):

$$s_n^{(CC_i)} \leq s_{CC_i}^{(CC_j)}, \qquad \text{and}, \qquad s_{CC_j}^{(d)} \leq s_{CC_i}^{(CC_j)}.$$

Implicit in this phase is that if (16c) is satisfied with equality in phase 2, only the design pertaining to cluster i or that pertaining to cluster j will be amenable to refinement.

This framework is summarized in Figure 6. In Section VII we will provide an instance in which this framework is used to design a networks with 160 nodes.

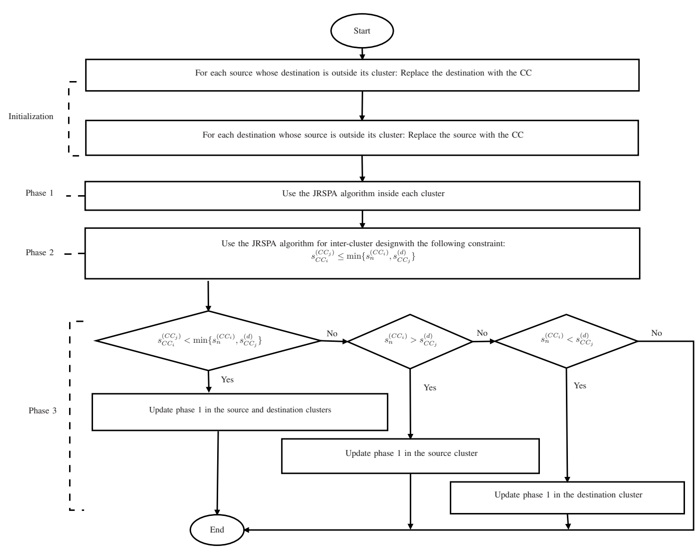


Fig. 6: Flowchart of the proposed framework for joint design of large clustered networks.

VII. SIMULATIONS

In this section, we assess the performance of the iterative algorithm and the proposed framework presented in Sections IV and VI. The optimization problems in this section are solved using the software package CVX [28] with an underlying MOSEK solver [29].

We consider a standard communication channel model with quasi-static frequency-flat Rayleigh fading subcarriers, log-normal shadowing, and path loss components. As such, the complex subcarrier gains can be expressed as $h_{\ell k} = \sqrt{\eta(\ell)\lambda_\ell} r_{\ell k}$, where $\eta(\cdot)$ is the path loss function. Shadowing is represented by λ_ℓ, which is log-normal distributed with 0 dB mean and standard deviation σ_s dB. Fading is represented by $r_{\ell k}$, which is complex Gaussian distributed with zero mean and unit variance. To simulate practical communication scenarios, we selected the distance values and the log-normal shadowing and path loss parameters corresponding to the urban macro-cell (UMa) scenario of IMT-Advanced document [30]. For that scenario, $\sigma_s = 6$ dB and the noise power, $\sigma^2 = -174$ dBm/Hz. Setting the carrier frequency to 2 GHz and the elevation of each device to 1.5 m, the path loss of the nonline-of-sight channel in this model is given by $\eta(\ell) = 10^{-18.66-40.32\log_{10}(d_\ell)}$, where d_ℓ is the length of link ℓ in meters.

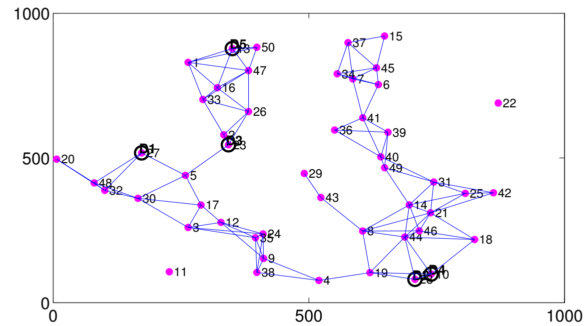


Fig. 7: Network topology

A. Performance Evaluation of the Proposed Scheme

In this section, we evaluate the performance of the scheme proposed in Section IV for a network instance with $N = 50$ nodes that are randomly distributed within a cell with a radius of 500 m. To maintain manageable computational cost, we allow two nodes to communicate only if the distance between them is smaller than 150 m. For the considered instance, the number of available links is $L = 208$, which are illustrated in Figure 7. Among the 50 nodes, 5 are randomly selected to act both as source and destination nodes, which are labelled as D_1 to D_5. The nodes are assumed to have identical power budgets, i.e, $P_n = P$, $n = 1, \cdots, 50$, and the available bandwidth is assumed to be 10 MHz, which is divided into 16 OFDM subcarriers.

The average sum rates yielded by the algorithm in Section IV for the values of P ranging from 0 to 30 dBm is depicted in Figure 8. For comparison, two baseline schemes are considered in this figure. The first is the joint optimization without power allocation, i.e., the output of the first stage in Algorithm 2 with fixed powers, and the second is the joint optimization without frequency-reuse in [17].

As can be seen from Figure 8, the sum rate yielded by the proposed scheme significantly outperforms the two baseline designs in which either frequency-reuse or power allocation is not considered. For instance, at $P = 15$ dBm, the proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

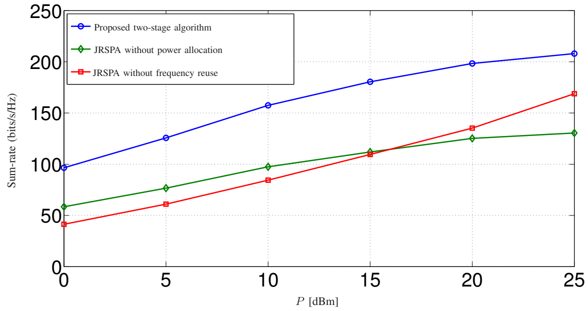


Fig. 8: Performance evaluation

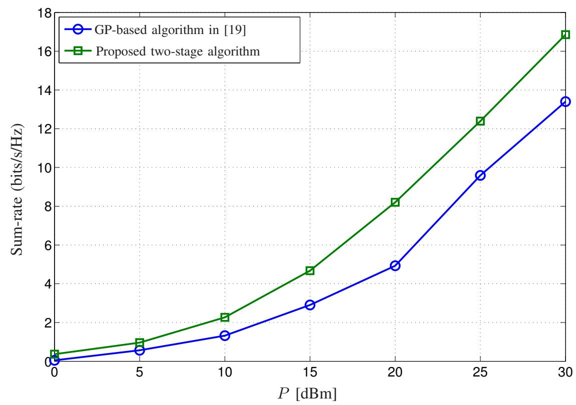
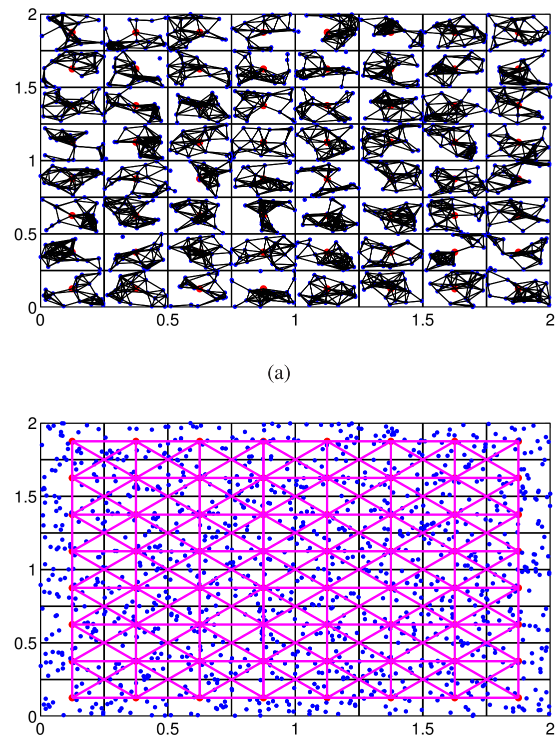


Fig. 9: Performance comparison



(a)

(b)

Fig. 10: Network topology and available links for communications (a) within clusters, (b) between CCs.

scheme yields a sum-rate advantage of 73% over the two baseline schemes.

B. Performance Comparison With the GP-Based Approach

In this section, we compare the performance of the algorithm proposed in Section IV with the one based on GP monomial approximation proposed in [19] for the network depicted in Figure 2. For this network, there are $N = 3$ nodes, $L = 6$ links and $K = 4$ subcarriers, each with a bandwidth of $W_0 = 200$ KHz. Two of the nodes wish to communicate with each other, with the third node potentially acting as a relay.

For the network considered in this example, we used the algorithm described in Section IV and the GP-based algorithm described in [19] to obtain the routes, schedules and power allocations and the average sum rates. The latter are depicted in Figure 9 for P ranging from 0 to 30 dBm.

From this figure, it can be seen that the proposed two-stage algorithm yields rates that are typically higher than those yielded by it GP-based counterpart. For instance at $P = 20$ dBm, the algorithm proposed herein yields an average sum rate 60% higher than that yielded by the GP-based algorithm. This phenomenon can be attributed to the ability of the algorithm proposed herein to use the gradient ascent approach to explore the feasible region for a good initial point, which contrasts the random initialization used in the scheme proposed in [19].

For convergence, we note that although the algorithm in [19] converges to a Karush-Kuhn-Tucker solution, it exhibits relatively slow convergence. In particular, for this network that algorithm converges within 180 iterations, which renders it impractical for designing larger networks. In contrast, the algorithm proposed herein converges within 10 iterations only.

C. Performance Evaluation of the Proposed Framework

In this section we evaluate the performance of the two-tier framework presented in Section VI. We considered an area of 2 Km^2 which is divided into 64 clusters as shown in Figures 10(a) and 10(b). Each cluster has 20 users whose locations are randomly chosen from the uniform distribution and there is a cluster controller at the center of each cluster. Hence, the total number of nodes in this network is 1344. In this network we assume that there are $K = 8$ subcarriers and 100 randomly chosen source-destination pairs (S_i, D_i), $i = 1, \ldots, 100$. To facilitate the design, we ignore intra-cluster and inter-cluster links greater than 100 and 400 meters, respectively. The remaining links available for communication are depicted in Figure 10(a) for communication between nodes within the clusters and in Figure 10(b) for communication between the CCs in the network. There are, on average, 176 links available for communication within each cluster.

To investigate the performance of the two-tier framework of Section VI in this network, we compare the average sum rate that it yields by optimizing the routes, schedules and power allocations with the average sum rate yielded when the nodes are restricted to have equal powers. In both cases the CCs are assumed to have identical power budgets of $P = 30$ dBm. For this case, the average sum rates yielded by the two-tier

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2816817, IEEE Access

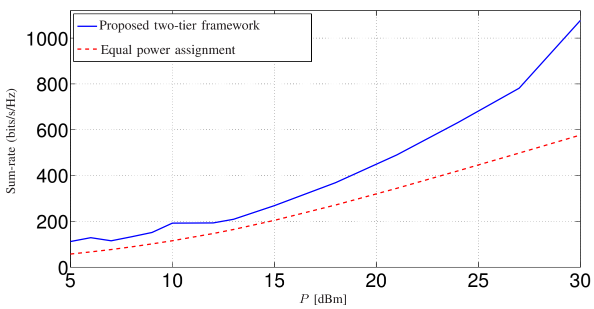


Fig. 11: Performance comparison of the proposed framework.

algorithm are depicted in Figure 11 for node budgets ranging from 5 to 30 dBm.

From Figure 11, it can be seen that the performance of the two-tier framework is significantly superior to the one with fixed power allocations. For instance, for an average sum rate of 400 bits/s/Hz, the two-tier approach has a power advantage of 5 dBm, and this advantage is larger for higher rates.

It is worth emphasizing that while the complexity of the algorithms in [19] allow the joint design of networks with up to 10 nodes, the two-tier algorithm in Section VI allows the joint design of significantly larger networks. In fact, our numerical evaluations suggest that this algorithm can be used to design networks with tens of thousands of nodes.

VIII. Conclusion

In this paper we considered the joint optimization of the routes, subcarrier schedules, time-shares and power allocations in large scale D2D communication networks. We made two main contributions. In the first contribution, we developed an iterative approach in which the design problem is decomposed into two sub-problems: one for scheduling and the other for power allocation. The latter is non-convex and to deal with it, we developed a constraint splitting approach, whereby the problem is further split into effectively two convex problems. The approach proceeds by performing inner iterations over the convex problems and outer iterations over the scheduling and power allocation sub-problems. This iterative approach is capable of jointly designing networks with up to 100 nodes. In the second contribution, we developed a two-tier approach whereby the network is divided into a set of non-overlapping clusters, each with a controller that acts as a gateway for managing inter-cluster communications. The first tier of this approach deals with intra-cluster communications, whereas the second tier deals with inter-cluster communications, both using the iterative algorithm developed in the first contribution. In comparison with existing algorithms, the ones developed herein yield better performance and can be used to design larger networks with significantly lower computational complexity.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, “What will 5G be?,” *IEEE J. Select. Areas Commun.*, vol. 32, pp. 1065–1082, June 2014.

[2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” *IEEE Commun. Mag.*, vol. 52, pp. 74–80, Feb. 2014.

[3] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. Fettweis, “Relay-based deployment concepts for wireless and mobile broadband radio,” *IEEE Commun. Mag.*, vol. 42, pp. 80–89, Sept. 2004.

[4] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y.-D. Kim, E. Kim, and Y.-C. Cheong, “An overview of radio resource management in relay-enhanced OFDMA-based networks,” *IEEE Commun. Surv. Tutorial*, vol. 12, no. 3, pp. 422–438, 2010.

[5] M. Rahman and H. Yanikomeroglu, “Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination,” *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1414–1425, Apr. 2010.

[6] C.-B. Chae, I. Hwang, R. W. Heath Jr, and V. Tarokh, “Interference aware-coordinated beamforming in a multi-cell system,” *IEEE Trans. Wireless Commun.*, vol. 11, pp. 3692–3703, Aug. 2012.

[7] I. F. Akyildiz and X. Wang, “A survey on wireless mesh networks,” *IEEE Commun. Mag.*, vol. 43, pp. 23–30, Sept. 2005.

[8] M. N. Tehrani, M. Uysal, and H. Yanikomeroglu, “Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 86–92, 2014.

[9] K.-D. Lee and V. C. Leung, “Fair allocation of subcarrier and power in an ofdma wireless mesh network,” *IEEE J. Select. Areas Commun.*, vol. 24, pp. 2051–2060, Nov. 2006.

[10] M. Tao and Y. Liu, “A network flow approach to throughput maximization in cooperative OFDMA networks,” *IEEE Trans. Wireless Commun.*, vol. 12, pp. 1138–1148, Mar. 2013.

[11] L. Xiao, M. Johansson, and S. P. Boyd, “Simultaneous routing and resource allocation via dual decomposition,” *IEEE Trans. Commun.*, vol. 52, pp. 1136–1144, July 2004.

[12] S. Hayashi and Z.-Q. Luo, “Spectrum management for interference-limited multiuser communication systems,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 1153–1175, Mar. 2009.

[13] K. Kim, Y. Han, and S.-L. Kim, “Joint subcarrier and power allocation in uplink OFDMA systems,” *IEEE Commun. Lett.*, vol. 9, pp. 526–528, June 2005.

[14] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, “Downlink scheduling and resource allocation for OFDM systems,” *IEEE Trans. Wireless Commun.*, vol. 8, pp. 288–296, Jan. 2009.

[15] L. Venturino, N. Prasad, and X. Wang, “Coordinated scheduling and power allocation in downlink multicell OFDMA networks,” *Proc. IEEE Vehic. Tech. Conf. (VTC)*, vol. 58, pp. 2835–2848, Jan. 2009.

[16] K. Karakayali, J. Kang, M. Kodialam, and K. Balachandran, “Cross-layer optimization for OFDMA-based wireless mesh backhaul networks,” in *Proc. IEEE Wireless Commun. Ntwk Conf. (WCNC)*, pp. 276–281, Mar. 2007.

[17] R. Rashtchi, R. H. Gohary, and H. Yanikomeroglu, “Routing, scheduling and power allocation in generic OFDMA wireless networks: Optimal design and efficiently computable bounds,” *IEEE Trans. Wireless Commun.*, vol. 13, pp. 2034–2046, Apr. 2014.

[18] H. Inaltekin and S. V. Hanly, “Optimality of binary power control for the single cell uplink,” *IEEE Trans. Inf. Theory*, vol. 58, pp. 6484–6498, Oct. 2012.

[19] R. Rashtchi, R. H. Gohary, and H. Yanikomeroglu, “Generalized cross-layer designs for generic half-duplex multicarrier wireless networks with frequency-reuse,” *IEEE Trans. Wireless Commun.*, vol. 15, pp. 458–471, Jan. 2016.

[20] A. Ahmad, S. Ahmad, M. H. Rehmani, and N. U. Hassan, “A survey on radio resource allocation in cognitive radio sensor networks,” *IEEE Commun. Surv. Tutorial*, vol. 17, pp. 888–917, Feb. 2015.

[21] M. Younis, K. Akkaya, and A. Kunjithapatham, “Optimization of task allocation in a cluster-based sensor network,” *Proc. IEEE Int. Symp. Computers Commun. (ISCC)*, pp. 329–334, Sept. 2003.

[22] A. Abdelnasser, E. Hossain, and D. I. Kim, “Clustering and resource allocation for dense femtocells in a two-tier cellular OFDMA network,” *IEEE Trans. Wireless Commun.*, vol. 13, pp. 1628–1641, Jan. 2014.

[23] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neil, and D. Julian, “Power control by geometric programming,” *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2640–2650, July 2007.

[24] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, “A tutorial on geometric programming,” *Optimization and Engineering*, vol. 8, pp. 67–127, Mar. 2007.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[26] G. Strang, *Introduction to Linear Algebra*. Wellesley, MA: Wellesley-Cambridge Press, 2009.
[27] P. Lin, J. Zhang, Y. Chen, and Q. Zhang, “Macro-femto heterogeneous network deployment and management: from business models to technical solutions,” *IEEE Wireless Commun.*, vol. 18, pp. 64–70, June 2011.
[28] M. Grant and S. Boyd, *CVX: MATLAB Software for Disciplined Convex Programming, version 1.21*, Jan. 2011. Available at: http://cvxr.com/cvx.
[29] MOSEK Apps., *The MOSEK Optimization Toolbox for MATLAB Manual, Version 6.0*, 2012. Available at: http://www.docs.mosek.com.
[30] Int. Telecommun. Union (ITU), *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced*. ITU-R: TR M.21351, Dec. 2009. Available at: http://www.itu.int/pub/R-REP-M.2135-1-2009.